Robust Background Removal in 4D Studio Images

Corina Blajovici
Babeş-Bolyai University
Cluj-Napoca, Romania
Email: blajovici@cs.ubbcluj.ro

Zsolt Jankó
Computer and Automation Research Institute
Budapest, Hungary
Email: janko@sztaki.hu

Dmitry Chetverikov
Computer and Automation Research Institute
Budapest, Hungary
Email: csetverikov@sztaki.hu

***Abstract*—In this paper we discuss background removal techniques developed for the 4D Reconstruction Studio at MTA SZTAKI. The 4D Studio enables creation of dynamic 3D models of real objects and actors. Robust foreground segmentation is a key element for 4D reconstruction, since the visual quality of the 3D model highly depends on the precision of the extracted silhouette. We present our novel solution for background removal based on exploiting the background colour in a robust manner. We also perform an analysis of shadow detection methods using different colour spaces, with the motivation of determining which colour representation is more suitable for shadow detection in a 4D Studio.**

I. Introduction

A 4D reconstruction studio is a room with uniform background, usually green or blue, equipped with lighting sources, multiple calibrated and synchronised video cameras and appropriate computing facilities. The main objective of a 4D reconstruction studio is to automatically build dynamic 3D models of real objects and actors, from sequences of simultaneous images taken from multiple viewpoints. The 4D studio at MTA SZTAKI [1], [2] is, to the best of our knowledge, the first 4D reconstruction studio in Eastern Europe. A brief survey of other advanced 4D reconstruction studios is presented in [1].

Image segmentation is a critical step in the 4D studio reconstruction pipeline, as it to a great extent determines the visual quality of the 3D model. As discussed in [3], [4], a two-phase strategy for image segmentation is proposed. In the first phase, the foreground silhouette is obtained by background subtraction using spherical colour representation. In the second phase, a further post-processing of the foreground image is proposed, aimed at detecting and removing shadows and extracting green regions belonging to the background. Our work is focused on background removal methods that are more robust and thus less sensitive to colour changes due to different clothing and illumination variations.

The main contribution of this work is twofold. First, we present a novel method for foreground segmentation based on exploiting the background colour in a robust manner. Second, we analyse how using different colour spaces for the shadow model can improve shadow detection in 4D studio images. We perform a comparative evaluation of different colour representations and discuss our results.

II. Related Work

Foreground segmentation is a key element in many computer vision applications such as 3D reconstruction from silhouettes, object tracking and vision-based motion capture systems. Background subtraction is one of the most common approaches used for separating moving foreground objects from the scene. In specific environments such as 4D studios, simpler methods have commonly been preferred due to their simplicity and efficiency for real-time processing. For their real-time 3D-modeling system, Petit et al. [5] use background subtraction in YUV colour space, based on a combination of a Gaussian model for the chromatic information and an interval model for the intensity information. Vlasic et al. [6] use in a multi-view reconstruction studio a combination of background subtraction and chroma-keying to obtain the foreground silhouettes.

Foreground objects obtained by different segmentation methods might still be noisy due to shadows that might be misclassified as foreground. A recent review published by Sanin et al. [7] describes methods for shadow detection in image sequences developed over the past decade. Most shadow detection approaches are based on the assumption that pixels in the shadow region are darker but have the same chromaticity. Various colour representations have been exploited for shadow detection (HSV [8], the normalised RGB [9], c1c2c3 [10], YUV [11]). Petit et al. [5] use for shadow detection in their real-time 3D reconstruction studio the method based on the approach of Horprasert et al. [12]. The shadow removal method developed for the 4D studio at MTA SZTAKI starts by detecting the ground region of the studio, using a simple single seeded region growing algorithm and then applying shadow detection inside this area [3], [4].

Some studies have been proposed in literature to evaluate the efficiency of colour representation for shadow detection. Benedek et al. [13] proposed a framework for evaluating the use of colour space for shadow detection in video sequences. They compared HSV, RGB, c1c2c3, CIE L*u*v colour spaces in simple indoor environments, highways with dark shadows and outdoor complex scenes with difficult illumination. For all these cases, they reported that CIE L*u*v was the most effective. Our study on colour space selection is targeted at specific environments such as the 4D studio.

III. Using Background Colour Information

In a dedicated environment such as the 4D studio, the background colour is essential information for segmentation in order to obtain high-quality, clean silhouettes. Since the studio is equipped with light sources and has a massive firm steel frame, the background is not uniformly green. Moreover, due to limitations of lighting, there exist darker regions in the background, where the green colour cannot be easily identified. Therefore, segmentation based only on the green colour to

obtain the foreground silhouette is not sufficient. However, the background colour can be used for post-processing in order to reduce green artefacts that are not removed by background subtraction.

There are a number of factors that make the task of detecting such misclassified green regions more difficult. Reflections of the green background on the foreground objects change the colour of the borders in areas where the colour is brighter or tends to be more bluish. Motion blur resulting from fast activities in the scene also introduces ambiguities between background and foreground.

In order to exploit the green colour information in a more robust manner, less sensitive to lighting, we define the green factors as follows:

$$\varphi_1 = \frac{I^G}{I^R}, \;\; \varphi_2 = \frac{I^G}{I^B}. \tag{1}$$

Here I^k, $k = R, G, B$ are the red, green and blue channels for each pixel in a given image I.

Using the green factors, we select a set of candidate pixels under the assumption that their green component is larger than the blue and the red one, based on the condition: $\varphi_i^B > \alpha_i \wedge \varphi_i^F > \alpha_i$, where $i = 1, 2$; $\varphi_i^B$and φ_i^F are the green factors for a given pixel in the background image and foreground image, respectively; α_i are thresholds set manually to 1.1 throughout all our tests. Threshold selection is important since some parts of the foreground, such as green reflections on the actor's cloths may also satisfy these conditions. For each pixel, the green factors are computed by taking the average of the green factors over a 5×5 window.

In order to make this result less sensitive to similarities between background and foreground, we further refine it by comparing the greenness of the candidate pixels in the background image and foreground image. If the green factors of a candidate pixel in the background are similar to the ones in the foreground, then the pixel belongs to the background, otherwise it belongs to the foreground. With this comparison the method is more reliable, as it can distinguish between green colour variations. In addition, rather than setting a global threshold for this comparison, we make it dependent on the background colour. This again improves robustness with respect to green reflections or other similarities in colour between the background and foreground. Thus, we decide that a candidate pixel belongs to the background if the following conditions are satisfied:

$$(1 - t \cdot \varphi_i^B) < \frac{\varphi_i^B}{\varphi_i^F} < (1 + t \cdot \varphi_i^B), \quad i = 1, 2. \tag{2}$$

Here $\varphi_i^B$and φ_i^F are the green factors of a given pixel in the background and the foreground image, respectively. The threshold t represents the tolerance to green and is usually set to $0.2 - 0.3$, depending on the greenness of the foreground. If parts of the foreground are green, selecting a smaller threshold value makes the algorithm more restrictive, assuring that parts of the foreground are not misclassified as background. Moreover, since the method does not make use of a global threshold, it can better handle situations when green reflection and other green parts add ambiguities in deciding whether a pixel belongs to the background or not.

IV. Shadow Detection

Shadow effects can be described by the shadow factor:

$$c^k = \frac{F^k(x)}{B^k(x)}, \;\; k = R, G, B. \tag{3}$$

Here F^k and B^k are the red, green and blue components of a shadow pixel in the frame F and the background image B, respectively. The shadow factor $c^k \in [0; 1]$ since shaded pixels are darker than the background.

This constant ratio rule is further exploited in a number of colour spaces. Shadow models based on colour representations such as HSV, YUV, CIE L*u*v have been used since they provide a natural separation into chromatic and luminance components. Photometric colour invariants, such as the normalised RGB, c1c2c3, l1l2l3 have also been exploited since they are invariant to changes in imaging conditions. In [13] it is shown that the constant ratio rule is a good approximation for the luminance components. In addition, although a surface becomes darker due to shadow, it preserves its chromaticity.

We evaluate shadow models in the RGB, HSV, c1c2c3 and normalized RGB colour representations. Similarly to [3], [4], shadow detection is applied only on the ground area. This improves the accuracy and efficiency of the overall process.

We choose the shadow model based on RGB colour space presented in [3], [4]. This is based on the assumption that in case of real shadows, the shadow factor must be the same at each channel. For HSV colour space we choose the method presented in [8]. Here the shadow model is formulated based on the assumption that when illumination changes, the value component changes, but hue and saturation changes are small. The reader is referred to [3], [4] for details on implementation of these two approaches.

Among the different photometric invariant colour spaces, we adopt the c1c2c3 and the normalized RGB representations. Using just a chromatic model and not taking into consideration the luminance can result in limitations in accuracy, especially for bright and dark objects, or objects similar in colour to the background. Therefore, we select as candidate shadow pixels, the pixels where the brightness is smaller than in the background image, such that the shadow factor $c^k < \tau_c$, for each $k = R, G, B$. The threshold τ_c is set to 0.9 to make the method more sensitive to changes in shading. Additionally, dark pixels are excluded from the analysis. Shadow factors are computed for each pixel over a 5×5 window, in order to make the method less sensitive to noise.

Shadow is then detected by making the assumption that the chromaticity components differ slightly in the shadow pixels. In c1c2c3 colour space, shadow pixels are selected based on following condition, similarly to [10]:

$$|c_1^F - c_1^B| < \tau \wedge |c_2^F - c_2^B| < \tau \wedge |c_3^F - c_3^B| < \tau, \tag{4}$$

where c_i^F, c_i^B are the c1c2c3 components for the foreground image and background image, respectively, and τ is a threshold empirically chosen for defining the fluctuations in chromaticity.

In the normalised RGB representation, which is often selected because of its fast calculation, the shadow is detected when the following condition is satisfied [9]:

$$|r^F - r^B| < \tau_r \wedge |b^F - b^B| < \tau_b. \tag{5}$$

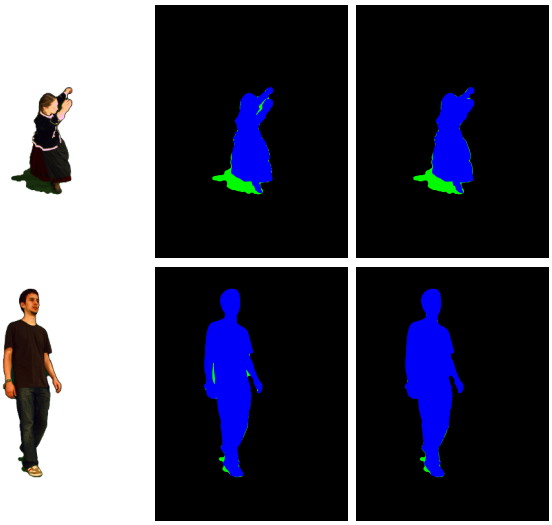


Fig. 1. Sample results of the proposed approach and the method in [3], [4]. Images in first column represent the silhouettes obtained by background subtraction. Second and third columns represent the result of the proposed approach, and the method in [3], [4], respectively. The foreground object is represented in blue, while the background regions in green.

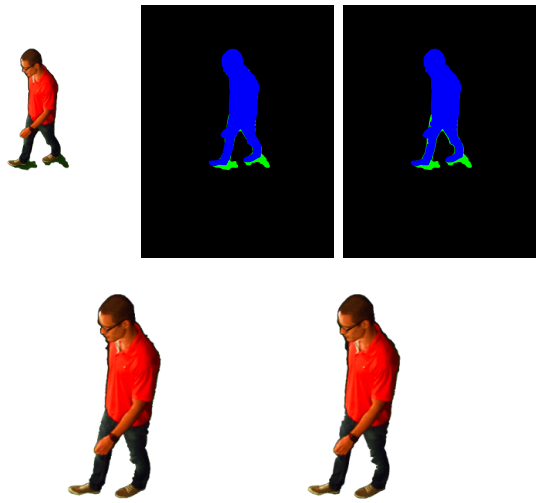


Fig. 2. First row: Input image and results obtained with the proposed approach and the method in [3], [4], respectively. The foreground object is shown in blue, the background regions in green. Second row: silhouettes obtained with both methods showing how the silhouette obtained by the proposed method is more accurate, even in the presence of green reflections on the actor's cloths.

Here r^F, r^B, b^F, b^B are the normalised RGB components of the image F and the background B. The differences are computed for a 5×5 window.

V. Test Results

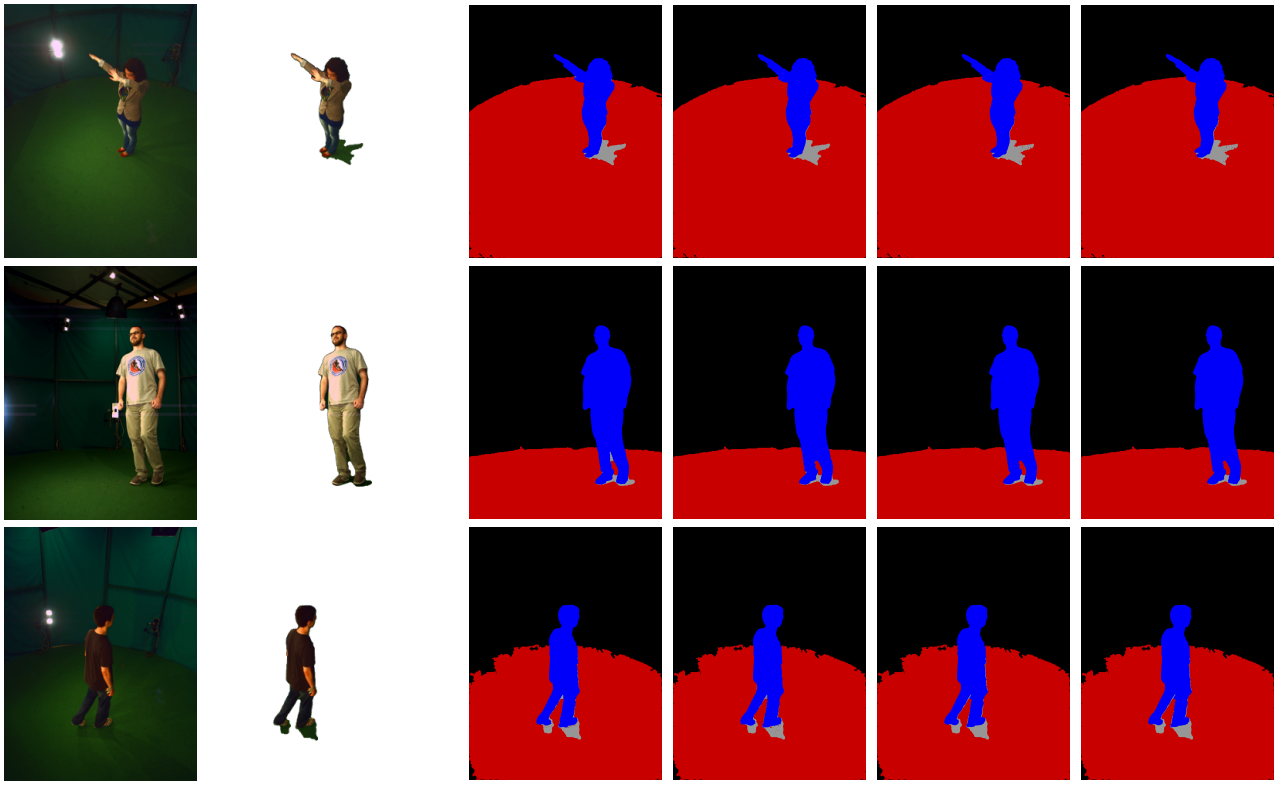
In order to test the post-processing algorithms, we have selected 8 sets of 13 images each, acquired by the 13 cameras of the 4D studio.

For evaluation of the proposed background removal using green colour, we have compared it with the approach described in [3], [4]. The results are shown in figure 1. The green colour detector proposed in [3], [4] is more sensitive to lighting conditions and to similarities between background and foreground, and needs tuning of parameters for some of the tested sequences, in order to optimise the visual appearance of the silhouette. On the other hand, the results obtained with the proposed approach have been generated using the thresholds $t = 2.5, \alpha_1 = 1.1, \alpha_2 = 1.1$, which are the same throughout all the tests. In addition, the proposed method performs better in darker green areas, since it depends on the background colour at each pixel, rather than on global thresholds such as the simple green detector. Our approach is more robust to green reflections of the background that may appear on the actor's cloths, as illustrated in figure 2. The method in [3], [4] removes such regions, visually distorting the contour of the foreground object shape.

Since shadows casted on the ground are also green, the proposed approach successfully removes such green shadow regions in areas where the shadows are not so strong. As illustrated in figure 3, shadow removal is still necessary. Figure 4 gives examples of results obtained with the four shadow detection methods. Note that all the methods show promising results. The shadow detection based on RGB colour space performs slightly better as it removes the dark regions between the legs.

Fig. 3. Result of the proposed green colour detector showing that shadow removal is still necessary. First image represents the input silhouette obtained by background subtraction. Second image represents the result obtained by the proposed green detector, while third image shows the result of the shadow detection method based on RGB [3], [4].

These methods have a few limitations, and we discuss them in the specific context of the 4D studio. Due to the assumption that the chromaticity is preserved in the shadow area, they are sensitive to colour similarities between foreground object and background. However, the shadow detection is applied only in the ground region, which improves the accuracy of the detection with respect to such similarities. In addition, all the methods are sensitive to dark regions, such as black clothing or dark hair, which tend to mislead the detection process. Different viewing angle provide different illumination conditions, and some sequences have darker frames with strong shadows created on the ground. For such sequences, results were obtained with different settings for the thresholds extracting the dark pixels, in order to improve the accuracy of the segmentation. However the thresholds are set once for the entire sequence, since it is not possible to change them for individual frames in the 4D studio pipeline. Figure 4 also gives an example of shadow detection for darker frames with dark clothing. Results by HSV and RGB are noisier in these cases, as compared to the normalised RGB and c1c2c3.



(a) Input Images (b) RGB (c) HSV (d) c1c2c3 (e) Normalised RGB

Fig. 4. Sample results of the four shadow detection methods. The foreground object is shown in blue, the shadow in grey, the ground region in red.

VI. Conclusion

In this paper, we introduced a novel technique for background removal based on green colour, aimed at improving foreground segmentation in 4D studio images. We have shown the robustness of the approach in various imaging conditions such as different clothing and illumination variations. In addition, since shadows casted on the ground might not be completely removed, we have analysed shadow removal using four colour representations, namely RGB, HSV, c1c2c3 and normalised RGB.

Our experience with the four methods for shadow removal can be summarised as follows. The shadow detection method based on RGB [3], [4] eliminates dark shadows better, but it is more sensitive to similarities between background and foreground than the other methods. In the HSV colour space, the shadow detection is more sensitive to dark areas, but because luminance is separated from chrominance, the detection is less sensitive to similarities. c1c2c3 gives better results than the normalised RGB, but both leave some small regions for strong shadows in dark areas and are sensitive to similarities in chromaticity.

References

[1] Z. Jankó, D. Csetverikov, and J. Hapák, "4D reconstruction studio: Creating dynamic 3D models of moving actors," in *Proc. 6th Hungarian Conf. Comput. Graph. and Geometry*, 2012, pp. 1–7.

[2] J. Hapák, Z. Jankó, and D. Chetverikov, "Real-time 4D reconstruction of human motion," in *Proc. 7th Int. Conf. Articulated Motion and Deformable Objects*, ser. Lecture Notes in Computer Science, vol. 7378, 2012, pp. 250–259.

[3] C. Blajovici, D. Chetverikov, and Z. Jankó, "4D studio for future internet: Improving foreground-background segmentation," in *Proc. IEEE 3rd Int. Conf. Cognitive Infocommunications*, 2012, pp. 559–564.

[4] ——, "Enhanced object segmentation in a 4D studio," in *Proc. Conf. Hungarian Assoc. for Image Process. and Pattern Recognition*, 2013, pp. 42–56.

[5] B. Petit, J.-D. Lesage *et al.*, "Multi-camera real-time 3D modeling for telepresence and remote collaboration," *Int. J. Digital Multimedia Broadcast.*, vol. 2010, January 2010.

[6] D. Vlasic, P. Peers *et al.*, "Dynamic shape capture using multi-view photometric stereo," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 174:1–174:11, 2009.

[7] A. Sanin, C. Sanderson, and B. C. Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern Recognition*, vol. 45, no. 4, pp. 1684–1695, 2012.

[8] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with hsv color information," in *Proc. IEEE Intell. Transp. Syst.*, Aug. 2001, pp. 334–339.

[9] A. Cavallaro, E. Salvador, and T. Ebrahimi, "Detecting shadows in image sequences," in *Proc. 1st European Conf. Visual Media Production*, 2004, pp. 15–16.

[10] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Computer Vision and Image Understanding*, vol. 95, no. 2, pp. 238–259, Aug. 2004.

[11] N. Martel-Brisson and A. Zaccarin, "Moving cast shadow detection from a gaussian mixture shadow model," in *Proc. CVPR*, vol. 2, 2005, pp. 643–648.

[12] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. ICCV*, vol. 99, 1999, pp. 1–19.

[13] C. Benedek and T. Szirányi, "Study on color space selection for detecting cast shadows in video surveillance," *Int. J. Imaging Syst. Technol.*, vol. 17, no. 3, pp. 190–201, Oct. 2007.